

Netflix Siblings

Taylor Blair

1/26/2021

Background

My sister and I have a long history of watching Netflix.

There are some minor outliers in the datasets, namely when we use each others accounts.

If you would like to create your own version of this dataset, or find your Netflix history go to: <https://www.netflix.com/YourAccount>

From there, select the specific profile \Rightarrow Viewing Activity \Rightarrow Download All

Loading Libraries and Data

```
library(tidyverse)
library(openair)
library(readr)

my_history <- read_csv("~/Coding/R/sis-and-me-netflix/MyNetflixHistory.csv")
sibling_history <- read_csv("~/Coding/R/sis-and-me-netflix/SisNetflixHistory.csv")

sample_n(my_history, 4)
```

	Title	Date
	<chr>	<chr>
## 1	Riverdale: Season 3: Chapter Forty-Three: Outbreak	3/31/20
## 2	Parks and Recreation: Season 5: Two Parties	3/9/20
## 3	Gossip Girl: Season 1: School Lies	5/12/20
## 4	Grey's Anatomy: Season 8: Dark Was the Night	10/5/19

Data Prep

As can be seen above, the initial dataset contains only two variables.

Date fixing

```
my_history$Date <- as.Date(my_history$Date, "%m/%d/%y")
sibling_history$Date <- as.Date(sibling_history$Date, "%m/%d/%y")
```

Movie or TV series

```
my_history$type <- grepl(":", my_history$Title)
my_history$type <- my_history$type %>%
  replace(my_history$type==TRUE, "TV Series") %>%
  replace(my_history$type==FALSE, "Movie")

sibling_history$type <- grepl(":", sibling_history$Title)
sibling_history$type <- sibling_history$type %>%
  replace(sibling_history$type==TRUE, "TV Series") %>%
  replace(sibling_history$type==FALSE, "Movie")
```

Series, Season and episode

```
my_history <- my_history %>%
  separate(Title,
    c("Series", "Season", "Episode"),
    ": ")

sibling_history <- sibling_history %>%
  separate(Title,
    c("Series", "Season", "Episode"),
    ": ")
```

Day of Week

Using

```
day_of_week <- function(x){
  return (weekdays(x))
}

days <- c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday")

my_history$day <- sapply(my_history$Date, day_of_week)
my_history$day <- factor(my_history$day,
  levels = days)

sibling_history$day <- sapply(sibling_history$Date, day_of_week)
sibling_history$day <- factor(sibling_history$day,
  levels = days)
```

Outputted Tibble

```
sample_n(my_history, 4)
```

```
## # A tibble: 4 x 6
##   Series          Season Episode          Date      type      day
##   <chr>          <chr>   <chr>        <date>    <chr>    <fct>
## 1 Medical Police Season 1 Real Heavy Hitter 2020-05-30 TV Series Saturday
## 2 The Umbrella Academy Season 2 Üga for Üga 2020-08-21 TV Series Friday
## 3 The Office (U.S.) Season 1 The Alliance 2019-07-02 TV Series Tuesday
## 4 House of Cards Season 1 Chapter 5 2019-07-03 TV Series Wednesday
```

#Compare dataset

```
my_history$individual <- "Tay"
sibling_history$individual <- "Sis"

merged <- rbind(my_history, sibling_history)
```

Numerical analysis

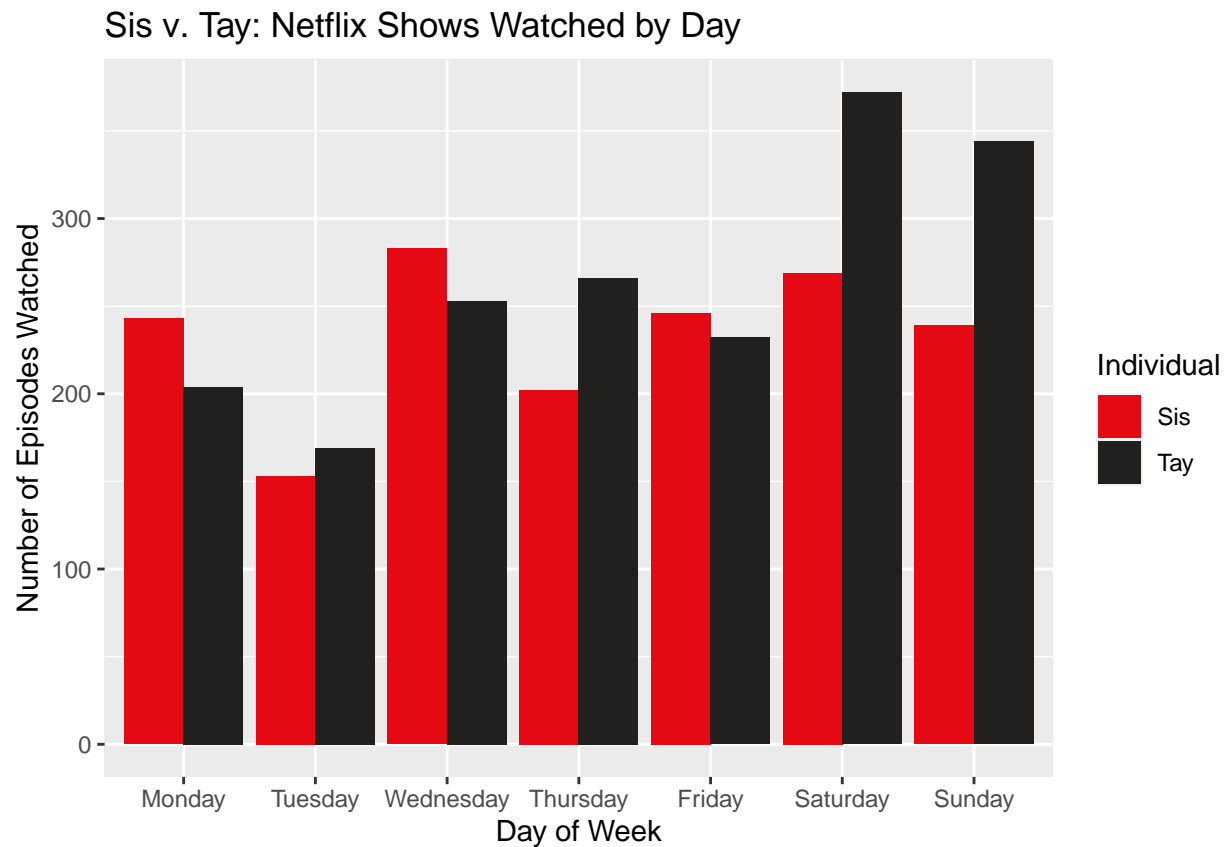
Binge Watching

Who hasn't binge watched Netflix before

Graphical Analysis

Looking at

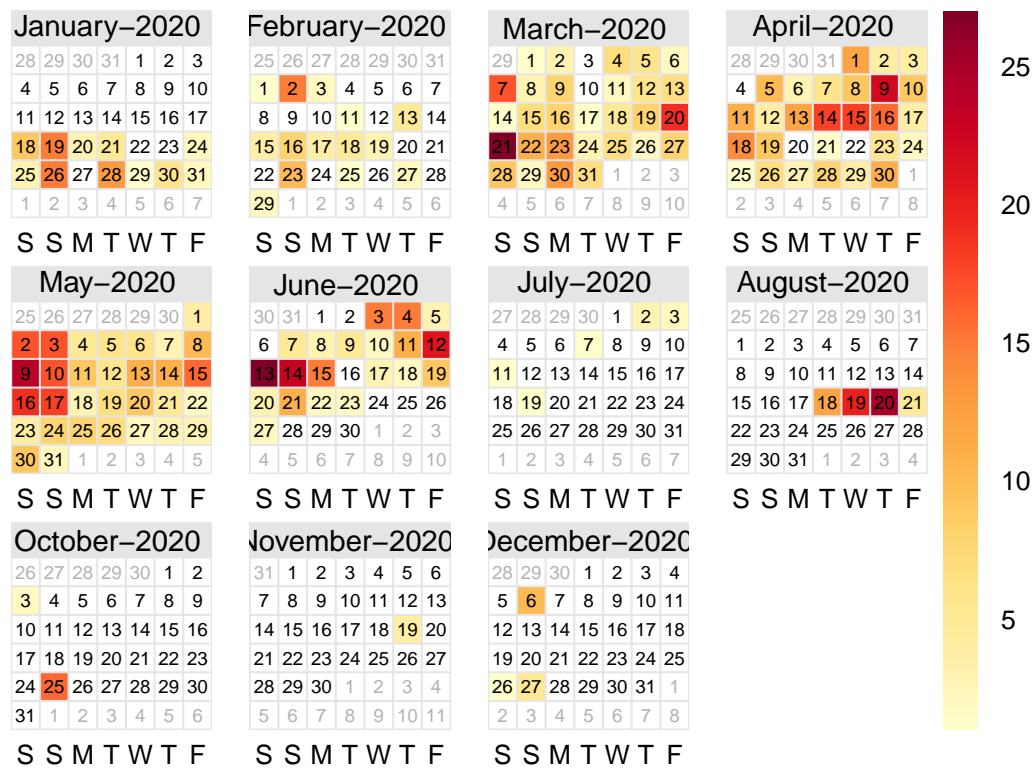
```
ggplot(merged, aes(x = day, fill=individual) ) +
  geom_bar(position = "dodge") +
  labs(title = "Sis v. Tay: Netflix Shows Watched by Day") +
  scale_fill_manual("Individual", values=c("#E50914", "#221F1F")) +
  xlab("Day of Week") +
  ylab("Number of Episodes Watched")
```



A Calander of Events

2020 was a unique year, with lockdown beginning in February...

```
my_history %>%
  group_by(Date) %>%
  summarise(count = n()) %>%
  rename(date=Date) %>%
  calendarPlot(pollutant = "count",
               year = 2020)
```



```
sibling_history %>%
  group_by(Date) %>%
  summarise(count = n()) %>%
  rename(date=Date) %>%
  calendarPlot(pollutant = "count",
               year = 2020,
               month = c(1:12))
```

